

Fisher-based thermodynamics for scale-invariant systems: Zipf's Law as an equilibrium state of a scale-free ideal gas

A Hernando¹, D Puigdomènech¹, D Villuendas² and C Vesperinas³

¹ Departament ECM, Facultat de Física, Universitat de Barcelona. Diagonal 647, 08028 Barcelona, Spain

² Departament FFN, Facultat de Física, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain

³ Sogeti España, WTCAP 2, Plaça de la Pau s/n, 08940 Cornellà, Spain

E-mail: alberto@ecm.ub.es, puigdomenech@ecm.ub.es, diego@ffn.ub.es and cristina.vesperinas@sogeti.com

Abstract. We present a thermodynamic formulation for scale-invariant systems based on the principle of extreme information. We create an analogy between these systems and the well-known thermodynamics of gases and fluids, and study as a compelling case the non-interacting system —the *scale-free ideal gas*— presenting some empirical evidences of electoral results, city population and total cites of Physics journals that confirm its existence. The empirical class of universality known as Zipf's law is derived from first principles: we show that this special class of power law can be understood as the density distribution of an equilibrium state of the scale-free ideal gas, whereas power laws of different exponent arise from equilibrium and non-equilibrium states. We also predict the appearance of the log-normal distribution as the equilibrium density of a harmonically constrained system, and finally derive an equivalent microscopic description of these systems.

PACS numbers: 89.70.Cf, 05.90.+m, 89.75.Da

1. Introduction

The study of scale-invariant phenomena has unravelled interesting and somewhat unexpected behaviours in systems belonging to disciplines of different nature, from physical and biological to technological and social sciences [1]. Indeed, empirical data from percolation theory and nuclear multifragmentation [2] reflect scale-invariant behaviour, and so do the abundances of genes in various organisms and tissues [3], the frequency of words in natural languages [4], scientific collaboration networks [5], the Internet traffic [6], Linux packages links [7], as well as electoral results [8], urban agglomerations [9] and firm sizes all over the world [10].

The common feature in these systems is the lack of a characteristic size, length or frequency for an observable k at study. This lack generally leads to a power law distribution $p(k)$, valid in most of the domain of definition of k ,

$$p(k) \sim 1/k^{1+\gamma}, \quad (1)$$

with $\gamma \geq 0$. Special attention has been paid to the class of universality defined by $\gamma = 1$, which corresponds to Zipf's law in the cumulative distribution or the rank-size distribution [2, 3, 4, 6, 7, 9, 10, 11]. Recently, Maillart et al. [7] have studied the evolution of the number of links to open source software projects in Linux packages, and have found that the link distribution follows Zipf's law as a consequence of stochastic proportional growth. In its simplest formulation, the stochastic proportional growth model, or namely the geometric Brownian motion, assumes the growth of an element of the system to be proportional to its size k , and to be governed by a stochastic Wiener process. The class $\gamma = 1$ emerges from the condition of stationarity, i.e., when the system reaches a dynamic equilibrium [11].

There is a variety of models arising in different fields that yield Zipf's law and other power laws on a case-by-case basis [9, 11, 12]. In the context of complex networks, proportional growth processes known as preferential attachment [6] and competitive cluster growth [13] have been used to explain many of the properties of natural networks, from social to biological. The emergence of power laws in all these models is explained by W. J. Redd and B. D. Hughes [14], which have shown analytically that models based on stochastic processes with exponential growth—as the geometric Brownian motion, discrete multiplicative process, the birth-and-death process, or the Galton-Watson branching process—generate power laws in one of both tails of the statistical distributions. However, in spite of the success of these models, the intrinsic complexity involved makes the study at a macroscopic level difficult since a general formulation of the thermodynamics of scale-invariant physics is not established yet.

Frieden et al. [15] have shown that equilibrium and non-equilibrium thermodynamics can be derived from the principle of extreme Fisher information. The information measure is done for the particular case of *translation families*, i.e., distribution functions whose form does not change under translational transformations. In this case, Fisher measure becomes *shift-invariant* [16], what yields most of the canonical Hamiltonians of theoretical physics [17]. Variations of the information measure lead to a Schrödinger equation [18] for the probability amplitude, where the ground state describes equilibrium physics and the excited states account for non-equilibrium physics. As for Hamiltonian systems [19], it has been recently shown that the principle of extreme physical information allows to describe the behaviour of complex systems, as the allometric or power laws found in biological sciences [20].

In this work we present a theoretical framework based on the principle of extreme physical information that aims to describe scale-invariant systems at a macroscopic level. We show that the thermodynamics for such systems can be formulated when the information measure is taken on distributions that do not change under *scale* transformations. We show that proportional growth is intrinsic to this symmetry, and the processes that describe Zipf's law, as the geometrical Brownian motion, are the equivalent microscopic description of these systems.

This work is organized as follows. In Sec. 2 we present the Fisher information measure for a scale-invariant system. In Sec. 3 we derive the equilibrium state and non-equilibrium states of the most general case of non-interacting scale-invariant system: the *scale-free ideal gas* (SFIG), and present some empirical evidences of its existence in electoral results and city population. In Sec. 4 we derive from first principles the special case of Zipf's law, what we call the *Zipf regime* of the SFIG, and study empirical data of the total number of cites of Physics journals to understand the conditions leading to its appearance. In Sec. 5 we constrain the system harmonically, finding that the equilibrium density follows a log-normal distribution. In Sec. 6 we derive from the SFIG the microscopic stochastic equation of motion, showing that the system can be described by geometrical Brownian walkers. Finally, in Sec. 7 we summarize our results and discuss some aspects of our work. In the Appendix we derive from the Fisher information the equations of the well-known translational-invariant ideal gas, which we use as analogy in the derivation of the SFIG.

2. The principle of extreme information for a scale-invariant system

The Fisher information measure I for a system of N elements, described by a set of coordinates \mathbf{q} and physical parameters θ , has the form [17]

$$I(F) = \int d\mathbf{q} F(\mathbf{q}|\theta) \sum_{ij} c_{ij} \frac{\partial}{\partial \theta_i} \ln F(\mathbf{q}|\theta) \frac{\partial}{\partial \theta_j} \ln F(\mathbf{q}|\theta), \quad (2)$$

where $F(\mathbf{q}|\theta)$ is the density distribution in configuration space (\mathbf{q}) conditioned by the physical parameters (θ). The constants c_{ij} account for dimensionality, and take the form $c_{ij} = c_i \delta_{ij}$ if q_i and q_j are uncorrelated. Following the principle of extreme information (PEI), the state of the system extremizes I subject to prior conditions, as the normalization of F or any constraint on the mean value of an observable $\langle A_i \rangle$. The PEI is then written as a variation problem of the form

$$\delta \left\{ I(F) - \sum_i \mu_i \langle A_i \rangle \right\} = 0, \quad (3)$$

where μ_i are the Lagrange multipliers. In the Appendix, we derive from the PEI the density distribution in configuration space and the entropy equation of state for the well-known translational invariant ideal gas (IG) [21]. In analogy with this derivation, we follow here the same steps to obtain the SFIG density distributions and entropy equations of state.

We consider a one-dimensional system with dynamical coordinates $\mathbf{q} = (k, v)$ where $dk/dt = v$. We define k as a discrete variable, i.e. $k = k_1, k_2, \dots, k_M$, where $k_i = i\Delta k$ and M , assumed to be large, is the total number of bins of width Δk . In order to address the scale-invariance behaviour of k in the Fisher formulation, we

change to the new coordinates $u = \ln k$ and $w = du/dt$, and assume that u and w are canonical [22] and uncorrelated. This assumption leads to the proportional growth

$$dk/dt = v = kw. \quad (4)$$

For constant w this equation yields an exponential growth $k = k_0 e^{wt}$, which is a uniform linear motion for u : $u = wt + u_0$, with $u_0 = \ln k_0$ ‡. It is easy to check that the scale transformation $k' = k/\theta_k$ leaves invariant the coordinate w , whereas the coordinate u transforms translationally $u' = u - \Theta_k$, where $\Theta_k = \ln \theta_k$.

If physics does not depend on scale, i.e., the system is translationally invariant with respect to the coordinates u and w , the distribution of physical elements can be described by the monoparametric translation families $F(u, w | \Theta_k, \Theta_w) = f(u', w')$. By analogy with the IG, we define the SFIG as a system of N non-interacting elements for which the density distribution can be factorized as $f(u, w) = g(u)h(w)$. Taking into account that u and w are canonical and uncorrelated ($c_{ii} = c_i \neq 0$ and $c_{uw} = c_{wu} = 0$), and that the Jacobian for the change of variables is $dkdv = e^{2u} du dw$, the information measure $I = I_u + I_w$ can be obtained in the continuous limit as

$$\begin{aligned} I_u &= c_u \int du e^{2u} g(u) \left| \frac{\partial \ln g(u)}{\partial u} \right|^2 \\ I_w &= c_w \int dw h(w) \left| \frac{\partial \ln h(w)}{\partial w} \right|^2. \end{aligned} \quad (5)$$

The constraints to the given observables $\langle A_i \rangle$ in the extremization problem determine the behaviour of the system. In the next sections we study three different cases: the general case of the scale-free ideal gas —the step-by-step analogy of the ideal gas—, a un-constrained gas or what we call the *Zipf regime*, and the harmonically constrained gas.

3. The scale-free ideal gas

For the general case, in the extremization of Fisher information we constrain the normalization of $g(u)$ and $h(w)$ to the total number of particles N and to 1, respectively

$$\int du e^{2u} g(u) = N, \quad \int dw h(w) = 1. \quad (6)$$

In addition, we penalize infinite values for w with a constraint on the variance of $h(w)$ to a given measured value

$$\int dw h(w) (w - \bar{w})^2 = \sigma_w^2, \quad (7)$$

where \bar{w} is the average growth. The variation yields

$$\delta \left\{ c_u \int du e^{2u} g \left| \frac{\partial \ln g}{\partial u} \right|^2 + \mu \int du e^{2u} g \right\} = 0 \quad (8)$$

and

$$\delta \left\{ c_w \int dw h \left| \frac{\partial \ln h}{\partial w} \right|^2 + \lambda \int dw h (w - \bar{w})^2 + \nu \int dw h \right\} = 0, \quad (9)$$

‡ This exponential growth allows to recognize the systems that we study in this work at the macroscopic level with those studied in [14].

where μ , λ and ν are Lagrange multipliers. Introducing $g(u) = e^{-2u}\Psi^*(u)\Psi(u)$, and varying (8) with respect to Ψ^* leads to the Schrödinger equation

$$\left[-4\frac{\partial^2}{\partial u^2} + 4 + \mu'\right]\Psi(u) = 0, \quad (10)$$

where $\mu' = \mu/c_u$. Analogously to the IG, we impose solutions compatible with a finite normalization of g in the thermodynamic limit $N, \Omega \rightarrow \infty$ with $N/\Omega = \rho_0$ finite, where $\Omega = \ln(k_M/k_1) = \ln M$ is the volume in u space and ρ_0 is defined as the *bulk density*. Solutions compatible with the normalization of (6) are given by $\Psi(u) = A_\alpha e^{-\alpha u/2}$, where A_α is the normalization constant and $\alpha = \sqrt{4 + \mu'}$. In this general case, the density distribution as a function of k takes the form of a power law: $g_\alpha(\ln k) = A^2/k^{2+\alpha}$. The equilibrium is defined by the ground state solution, which correspond the lowest allowed value $\alpha = 0$. It can be show that it is just a uniform density distribution in u space at the bulk density: $g(u)e^{2u}du = N/\Omega du = \rho_0 du$.

Introducing $h(w) = \Phi^*(w)\Phi(w)$ and varying (9) with respect to Φ^* leads to the quantum harmonic oscillator equation [18]

$$\left[-4\frac{\partial^2}{\partial w^2} + \lambda'(w - \bar{w})^2 + \nu'\right]\Phi(w) = 0, \quad (11)$$

where $\lambda' = \lambda/c_w$ and $\nu' = \nu/c_w$. The equilibrium configuration corresponds to the ground state solution, which is now a Gaussian distribution. Using (7) to identify $|\lambda'|^{-1/2} = \sigma_w^2$ we get the Boltzmann distribution

$$h(w) = \frac{\exp[-(w - \bar{w})^2/2\sigma_w^2]}{\sqrt{2\pi}\sigma_w}. \quad (12)$$

The density distribution in configuration space $\tilde{f}(k, v)dkdv = f(u, w)e^{2u}dudw$ is then

$$\tilde{f}(k, v) = \frac{N}{\Omega k^2} \frac{\exp[-(v/k - \bar{w})^2/2\sigma_w^2]}{\sqrt{2\pi}\sigma_w}. \quad (13)$$

If we define $H = \Delta k^2/\Delta\tau$ as the elementary volume in phase space, where $\Delta\tau$ is the time element, the total number of microstates is $Z = N!H^N \prod_{i=1}^N f_1(k_i, v_i)$, where $f_1 = \tilde{f}/N$ is the monoparticular distribution function and $N!$ counts all possible permutations for distinguishable elements. The entropy equation of state $S = -\kappa \ln Z$ reads

$$S = N\kappa \left\{ \ln \frac{\Omega}{N} \frac{\sqrt{2\pi}\sigma_w}{H'} + \frac{3}{2} \right\}, \quad (14)$$

where κ is a constant that accounts for dimensionality and $H' = H/(k_M k_1) = H/(M\Delta k^2) = 1/(M\Delta\tau)$. Remarkably, this expression has the same form as the one-dimensional IG ($D = 1$ in (A.9)); instead of the thermodynamical variables (N, V, T) , here we deal with the variables (N, Ω, σ_w) , which make the entropy scale-invariant.

The total density distribution for k is obtained integrating for all v the density distribution in configuration space. Integrating (13) we get

$$\tilde{f}(k) = \int dv \tilde{f}(k, v) = \frac{N}{\Omega} \frac{1}{k} = \frac{\rho_0}{k}, \quad (15)$$

which corresponds for large N to an exponential rank-size distribution

$$k(r) = k_1 \exp \left[\Omega - \frac{r-1}{\rho_0} \right], \quad (16)$$

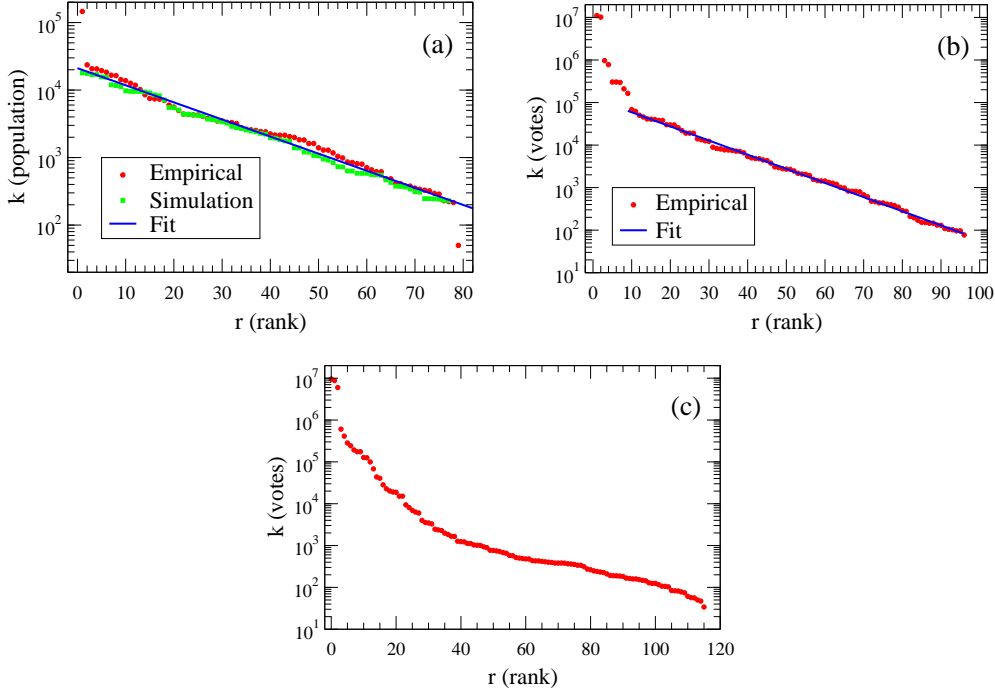


Figure 1. (colour on-line) **a**, rank-size distribution of the cities of the province of Huelva, Spain (2008), sorted from largest to smallest, compared with the result of a simulation with Brownian walkers (green squares). **b**, rank-plot of the 2008 General Elections results in Spain. **c**, rank-plot of the 2005 General Elections results in the United Kingdom. (red dots: empirical data; blue lines: linear fitting).

where r is the rank.

This behaviour, which corresponds to the class of universality $\gamma = 0$ in (1), has been empirically found by Costa Filho et al. [8] in the distribution of votes in the Brazilian electoral results. We have found such a behaviour in the city-size distribution of small regions and electoral results, like the province of Huelva (Spain) [23], and the 2008 Spanish General Elections results [24], respectively. We show in figure 1a and 1b their rank-size distribution in semi-logarithmic scale, where a straight line corresponds to a distribution of type (16). Most of the distribution can be linearly fitted, with a correlation coefficient of 0.994 and 0.998 respectively. From these fits we have obtained a bulk density of $\rho_0 = 0.058$ for the General Elections results, and in the case of Huelva $\rho_0 = 17.1$ ($N = 77$, $\Omega = 4.5$). Using historical data for the latter [23], we have used the backward differentiation formula to calculate the relative growth rate of the i -th city as

$$w_i = \frac{\ln k_i^{(2008)} - \ln k_i^{(2007)}}{\Delta t} \quad (17)$$

where $k_i^{(2007)}$ and $k_i^{(2008)}$ are the number of inhabitants of the i -th city in 2007 and 2008 respectively and $\Delta t = 1$ year. We have obtained $\bar{w} = 0.012 \text{ years}^{-1}$ and $\sigma_w = 0.032 \text{ years}^{-1}$. However, the regularities are not always obvious, as shown

for the most voted parties in Spain'08 or the whole distribution of the 2005 General Elections results in the United Kingdom [25] (figure 1c). In both cases, the competition between parties seems to play an important role, and the assumption of non-interacting elements can be unrealistic §.

4. The Zipf regime

In the previous subsection we considered that N/Ω remains finite even in the thermodynamic limit, i.e., the system reaches the bulk density ρ_0 . However, if $N/\Omega \rightarrow 0$ as $\Omega \rightarrow \infty$, i.e., the system is exposed to an empty infinite volume, the normalization can not be achieved and the constraint has to be removed ($\mu = 0$). We call this case the *Zipf regime*, in order to distinguish it from the general.

Considering only the k coordinate in the domain $[k_1, \infty)$, the information measure for the total density distribution $\tilde{f}(k)dk = f(u)e^u du = p(u)du$, reads

$$I_u = c_u \int du e^u f(u) \left| \frac{\partial \ln f(u)}{\partial u} \right|^2, \quad (18)$$

and the extremization problem

$$\delta \left\{ c_u \int du e^u f(u) \left| \frac{\partial \ln f(u)}{\partial u} \right|^2 \right\} = 0. \quad (19)$$

Introducing $f(u) = e^{-u} \Psi^*(u) \Psi(u)$, and varying with respect to Ψ^* leads to the Schrödinger equation

$$\left[-4 \frac{\partial^2}{\partial u^2} + 1 \right] \Psi(u) = 0. \quad (20)$$

Taking the boundary conditions $\lim_{u \rightarrow \infty} f(u) = 0$ and $f(u_1) = C$ where $u_1 = \ln k_1$ and C is a constant, the solution to the equation is $\Psi(u) = C' e^{-u/2}$, where $C' = \sqrt{C} e^{u_1}$. It can be shown that this is just an exponential decay in u space $f(u)e^u du = C e^{-u} du$. This solution leads to the total density distribution

$$\tilde{f}(k) = \frac{C^2}{k^2} \quad (21)$$

with $C^2 = N k_1$ for a normalized density. It corresponds in the continuous limit to a rank-size distribution of the type

$$k(r) = \frac{N k_1}{r}, \quad (22)$$

which is the Zipf's law (universal class $\gamma = 1$) of [2, 3, 4, 6, 7, 9, 10, 11]. This result is remarkable: for the first time *Zipf's law is derived from first principles*.

In figure 2 we show the known behaviour [11] of the rank size distribution of the top 100 largest cities of the United States [27], which shows an slope near -1 ($\gamma = 1$) in the logarithmic representation of the rank-plot.

The appearance of the bulk and the Zipf regime in a SFIG can be understood studying empirical data. We have studied the system formed by all Physics journals [28] ($N = 310$) using their total number of cites as coordinate k . If a journal receives more cites due to its popularity, it becomes even more popular and therefore it

§ The effects of interaction are studied in [26], where we go beyond the non-interacting system using a microscopic description based on complex networks.

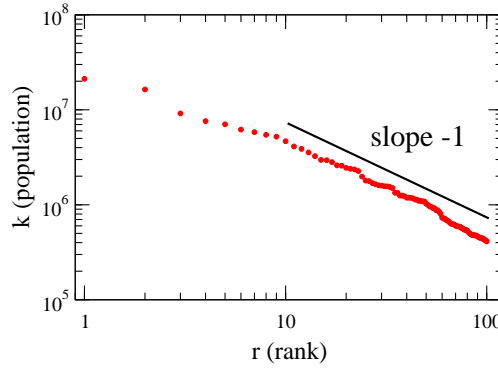


Figure 2. (colour on-line) Rank-plot of the 100 largest cities of the United States.

will receive more cites. Under such conditions proportional growth and scale invariance are expected. Since we consider all fields of Physics, correlation effects are much lower than only consider journals of an specific field, so the non-interacting approximation seems realistic in this case. In figure 3 we show the rank-plot of the number of cites of Physic journals, where we have found a slope near -1 for the most-cited journals in the logarithmic representation (figure 3a) and an slope near $+1$ for the less-cited journals (figure 3b). For the central part of the distribution bulk density reaches a value of $\rho_0 \sim 57$ (figure 3c).

This distribution shows an extraordinary symmetric behaviour under the change $k \rightarrow 1/k$ ($u \rightarrow -u$). We show in figure 4 the raw empirical data compared with the distribution obtained from the transformation $k' = c/k$ ($u' = -u + \ln c$), where $c = 3.3 \times 10^6$. The symmetry of this system is an important clue to understand both regimes, and represents a perfect example of the conditions needed to observe bulk and Zipf regimes in a non-interacting scale-invariant system. The main part of the density distribution reaches the bulk density obeying (15), whereas Zipf's law emerge at the edges, obeying (21): following the analogy with the physics of gases and fluids, we can think of the system as a drop, where the Zipf regime is the sign of a *surface* since it reproduces how the density exponentially falls from the bulk density to zero in u space when the system is exposed to an infinite empty volume. This effect is clearly visible in figure 5, where the empirical density distribution $p(u)du$ in u space is compared with the fitted density

$$p(u) = \begin{cases} \rho_a e^u & \text{if } u < u_a \\ \rho_0 & \text{if } u_a < u < u_b \\ \rho_b e^{-u} & \text{if } u_b > u \end{cases} \quad (23)$$

where $\rho_a = 0.1$, $\rho_0 = 57$, $\rho_b = 4.3 \times 10^5$, $u_a = 5.2$ and $u_b = 10$. These findings lead us to conclude that the system of Physics journals sorted by total number of cites is a perfect example of the scale-free ideal gas at equilibrium.

5. The harmonically constrained system

We now consider a system with a constraint in a given observable $\langle A \rangle$ which locally depends on k , $A = A(k)$. The second order Taylor expansion with respect to $u = \ln k$ near a minimum is written as $\tilde{A}(u) = A(e^u) \simeq A_0 + A_2/2(u - u_m)^2$, where A_0 , A_2

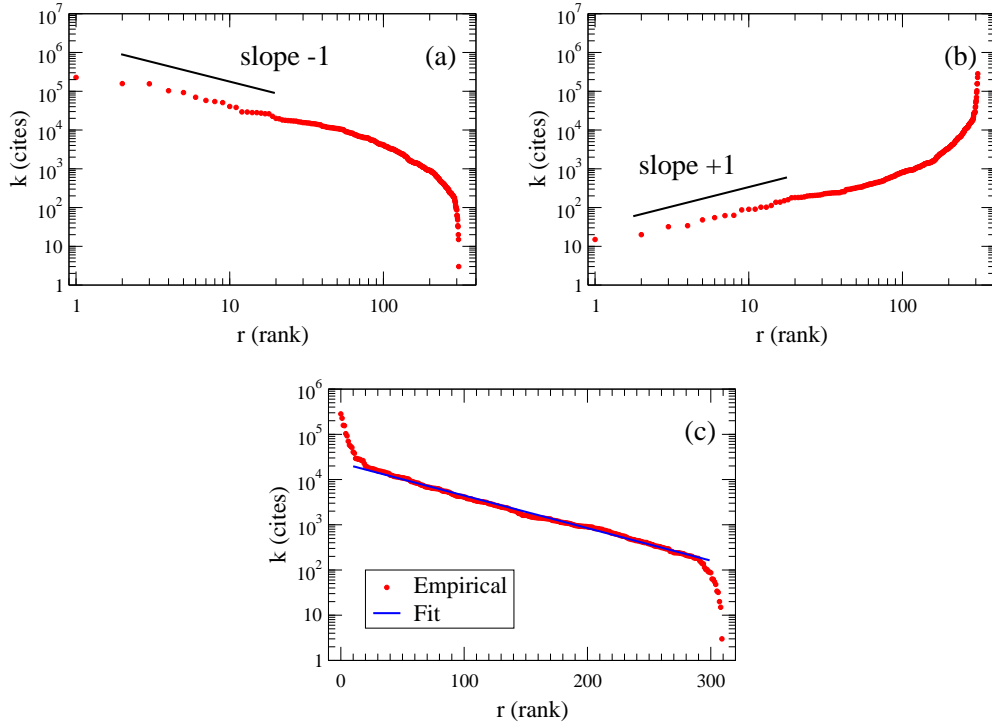


Figure 3. (colour on-line) **a**, rank-plot of the total number of cites of physics journal, from most-cited to less-cited, in logarithm scale. **b**, sorted from less-cited to most cited **c**, same as **a**, in semi-logarithm scale. (red dots: empirical data; blue line: linear fitting).

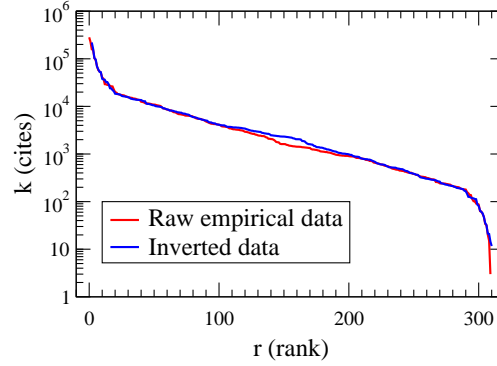


Figure 4. (colour online) Rank-plot of the total number of cites of Physics journal, from most-cited to less-cited, compared with the distribution obtained from the inverse transformation $k' = 3.3 \times 10^6/k$ where k is the number of cites.

and u_m are constants. Introducing this constraint and the normalization condition to the number of elements N of the total density distribution, the extremization problem

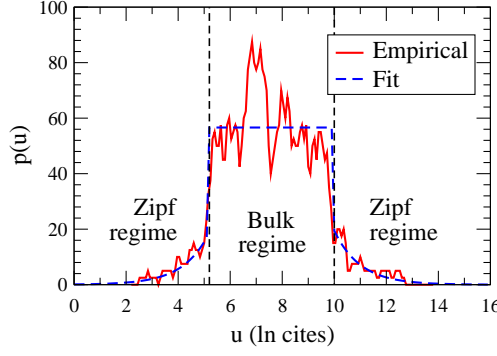


Figure 5. (colour online) Empirical density distribution in u space of the total number of cites of Physics journals, compared with (23). The bulk regime and the Zipf regime at the edges is clearly visible.

reads

$$\delta \left\{ c_u \int du e^u f \left| \frac{\partial \ln f}{\partial u} \right|^2 + \mu \int du e^u f + \lambda \int du e^u f \left[A_0 + \frac{1}{2} A_2 (u - u_m)^2 \right] \right\} = 0 \quad (24)$$

Introducing $f(u) = e^{-u} \Psi^*(u) \Psi(u)$, and varying with respect to Ψ^* leads to the quantum harmonic oscillator equation

$$\left[-4 \frac{\partial^2}{\partial u^2} + \lambda' (u - u_0)^2 + \nu' \right] \Psi(u) = 0, \quad (25)$$

where now we have defined $\lambda' = (\lambda A_2)/2c_u$ and $\nu' = (\nu + 1 + \lambda A_0)/c_u$. The ground state solution is a gaussian distribution, which now yields a total density distribution of the form of a log-normal distribution

$$\tilde{f}(k) = \frac{N}{k \sqrt{2\pi\sigma_u}} \exp \left(-\frac{(\ln k - u_m)^2}{2\sigma_u^2} \right), \quad (26)$$

with $|\lambda'|^{-1/2} = \sigma_u^2 = \langle A \rangle - A_0$. Note that if $A_0 = 0$, the constraint can be also understood as a constraint in the variance of u .

The log-normal distribution has been widely observed in a large number of scale-invariant systems [29]. In [30] S. Fortunato and C. Castellano found this behaviour in the electoral results of different countries and for different years. We can think of this constraint as the effect of polices or social factors: low popularity candidates are penalized since the party does not present them for the elections, and high popularity candidates are penalized by the competition in campaign. Both effects can be approximated to second order as a harmonic potential, however anharmonic effects are expected in a high order study.

Defining $H'' = 1/\Delta\tau$ and $k_m = \ln u_m$ being $k_m = m\Delta k$, the entropy equation of state reads in this case

$$S = N\kappa \left\{ \ln \frac{2\pi\sigma_u\sigma_w m^2}{NH''} + 2 \right\}, \quad (27)$$

which maintains scale invariance.

6. The microscopic description

The dynamics of the system can be microscopically described as a stochastic process using (4) and the density distribution (12). Treating w as a random variable, the stochastic equation of motion is written as a geometrical Brownian motion

$$dk = k\bar{w}dt + k\sigma_w dW, \quad (28)$$

where dW is a Wiener process. In the u space, this equation reads

$$du = \bar{w}dt + \sigma_w dW, \quad (29)$$

which describes the well-known Brownian motion. (28) exactly describes the dynamical condition found empirically in [7] and also the stochastic proportional growth model used in [11] to obtain Zipf's law. We can think of this sort of simulations as the equivalent of molecular dynamics simulations for gases and liquids [31].

Effectively, (29) implies that a uniform density in u space of N Brownian walkers moving in a fixed volume Ω —a model used in the literature to describe the IG [31]—describes the SFIG when we represent the system with the coordinates (k, v) . In figure 1a we show the rank-plot for a system of $N = 78$ geometrical Brownian walkers with $\sigma_w = 0.029$ in a volume of $\Omega = 4.5$ and $k_1 = 200$ in reduced units, which nearly describes the distribution of the population of the province of Huelva.

7. Summary and Discussion

We have shown that a thermodynamic description of scale-invariant systems can be formulated from the principle of extreme information, finding an analogy with the thermodynamics of gases and fluids. We have derived the density distribution in configuration space and the entropy equation of state of the scale-free ideal gas in the thermodynamic limit, and have found empirical evidences of its existence in city population, electoral results and cites to Physics journals. In this context, Zipf's law emerges naturally as the equilibrium density of the non-interacting system when the volume grows to infinity, what we call the Zipf regime. Using empirical data we have seen that this regime can be understood as the density fall of a surface between the bulk and an empty volume. We have also studied the effect of a harmonic constraint, finding that in this case the density of the system follows a log-normal distribution, which has been empirically observed in electoral results and in many other scale-invariant systems [29]. Finally we have shown with a simulation of city population that a geometrical Brownian motion can describe the system at a microscopic level.

It is well known that in real gases the most interesting situations emerge when interactions between particles become relevant, originating deviations from the equation of state of the IG, and making room for the appearance of, e.g., phase transitions [21]. Analogously, one should also expect this rich phenomenology to show up in scale-invariant real systems, which may explain deviations from Zipf's law in empirical distributions. A study beyond the ideal gas is in progress, and further results will be reported [26].

Acknowledgments

We would like to thank M. Barranco, R. Frieden, A. Plastino, and B. H. Soffer for useful discussions. This work has been partially performed under grant FIS2008-00421/FIS from DGI, Spain (FEDER).

Appendix A. The translational invariant ideal gas

In this appendix we derive from the principle of extreme information the density distribution in configuration space and the entropy equation of state of the translational invariant ideal gas (IG) [21]. The IG model describes non-interacting classical particles of mass m with coordinates $\mathbf{q} = (\mathbf{r}, \mathbf{p})$, where $m d\mathbf{r}/dt = \mathbf{p}$. We assume that these coordinates are canonical [22] and uncorrelated. This assumption is introduced in the information measure (2) as $c_{ij} = c_i \delta_{ij}$, where $c_i = c_r$ for space coordinates, $c_i = c_p$ for momentum coordinates, and δ_{ij} is the Kronecker delta. The density distribution can be factorized as $f(\mathbf{r}, \mathbf{p}) = \rho(\mathbf{r})\eta(\mathbf{p})$, and the information measure $I = I_r + I_p$ reads, if D is the dimension of the space

$$\begin{aligned} I_r &= c_r \int d^D \mathbf{r} \rho(\mathbf{r}) |\nabla_r \ln \rho(\mathbf{r})|^2 \\ I_p &= c_p \int d^D \mathbf{p} \eta(\mathbf{p}) |\nabla_p \ln \eta(\mathbf{p})|^2. \end{aligned} \quad (\text{A.1})$$

In the extremization of Fisher information we constrain the normalization of $\rho(\mathbf{r})$ and $\eta(\mathbf{p})$ to the total number of particles N and to 1, respectively

$$\int d^D \mathbf{r} \rho(\mathbf{r}) = N, \quad \int d^D \mathbf{p} \eta(\mathbf{p}) = 1. \quad (\text{A.2})$$

In addition, we penalize infinite values for the particle momentum with a constraint on the variance of $\eta(\mathbf{p})$ to a given measured value

$$\int d^D \mathbf{p} \eta(\mathbf{p}) (\mathbf{p} - \bar{\mathbf{p}})^2 = D \sigma_p^2, \quad (\text{A.3})$$

where $\bar{\mathbf{p}}$ is the mean value of \mathbf{p} . For each degree of freedom it is known from the Virial theorem that the variance is related to the temperature T as $\sigma_p^2 = mk_B T$, being k_B the Boltzmann factor. The variation yields

$$\delta \left\{ c_r \int d^D \mathbf{r} \rho |\nabla_r \ln \rho|^2 + \mu \int d^D \mathbf{r} \rho \right\} = 0 \quad (\text{A.4})$$

and

$$\delta \left\{ c_p \int d^D \mathbf{p} \eta |\nabla_p \ln \eta|^2 + \lambda \int d^D \mathbf{p} \eta (\mathbf{p} - \bar{\mathbf{p}})^2 + \nu \int d^D \mathbf{p} \eta \right\} = 0, \quad (\text{A.5})$$

where μ , λ and ν are Lagrange multipliers.

Introducing $\rho(\mathbf{r}) = \Psi^*(\mathbf{r})\Psi(\mathbf{r})$ and varying (A.4) with respect to Ψ^* leads to the Schrödinger equation [18]

$$[-4\nabla_r^2 + \mu'] \Psi(\mathbf{r}) = 0, \quad (\text{A.6})$$

where $\mu' = \mu/c_r$. To fix the boundary conditions, we first assume that the N particles are confined in a box of volume V , and next we take the thermodynamic limit (TL) $N, V \rightarrow \infty$ with N/V finite. The equilibrium state compatible with this limit corresponds to the ground state solution, which is the uniform density $\rho(\mathbf{r}) = N/V$.

Introducing $\eta(\mathbf{p}) = \Phi^*(\mathbf{p})\Phi(\mathbf{p})$ and varying (A.5) with respect to Φ^* leads to the quantum harmonic oscillator equation [18]

$$[-4\nabla_p^2 + \lambda'(\mathbf{p} - \bar{\mathbf{p}})^2 + \nu'] \Phi(\mathbf{p}) = 0, \quad (\text{A.7})$$

where $\lambda' = \lambda/c_p$ and $\nu' = \nu/c_p$. The equilibrium configuration corresponds to the ground state solution, which is now a gaussian distribution. Using (A.3) to identify

$|\lambda'|^{-1/2} = \sigma_p^2$ we get the Boltzmann distribution, which leads to a density distribution in configuration space of the form

$$f(\mathbf{r}, \mathbf{p}) = \frac{N}{V} \frac{\exp[-(\mathbf{p} - \bar{\mathbf{p}})^2/2\sigma_p^2]}{(2\pi\sigma_p^2)^{D/2}}. \quad (\text{A.8})$$

If H is the elementary volume in phase space, the total number of microstates is $Z = N!H^{DN} \prod_{i=1}^N f_1(\mathbf{r}_i, \mathbf{p}_i)$, where $f_1 = f/N$ is the monoparticular distribution and $N!$ counts all possible permutations for distinguishable particles. The entropy $S = -k_B \ln Z$ is written as

$$S = Nk_B \left\{ \ln \frac{V}{N} \left(\frac{2\pi\sigma_p^2}{H^2} \right)^{D/2} + \frac{2+D}{2} \right\}, \quad (\text{A.9})$$

where we have used the Stirling approximation for $N!$. This expression is in exact accordance with the known value of the entropy for the IG [21], which shows the predictive power of the Fisher formulation.

References

- [1] *Fractals in Physics*, edited by Aharony A and Feder J 1989 *Proc. Conf. in honor of Mandelbrot B B, Vence, France* (North Holland, Amsterdam).
- [2] Paech K, Bauer W and Pratt S 2007 *Phys. Rev. C* **76**, 054603; Campi X and Krivine H 2005 *Phys. Rev. C* **72**, 057602; Ma Y G *et al.* 2005 *Phys. Rev. C* **71**, 054606.
- [3] Furusawa C and Kaneko K 2003 *Phys. Rev. Lett.* **90**, 088102.
- [4] Zipf G K 1949 *Human Behavior and the Principle of Least Effort* (Addison-Wesley Press, Cambridge, Mass.); Kanter I and Kessler D A 1995 *Phys. Rev. Lett.* **74**, 4559.
- [5] Newman M E J 2001 *Phys. Rev. E* **64**, 016131.
- [6] Barabasi A L and Albert R 2002 *Rev. Mod. Phys.* **74**, 47
- [7] Maillart T, Sornette D, Spaeth S and von Krogh G 2008 *Phys. Rev. Lett.* **101**, 218701.
- [8] Costa Filho R N, Almeida M P, Andrade J S and Moreira J E 1999 *Phys. Rev. E* **60**, 1067.
- [9] Malacarne L C, Mendes R S and Lenzi E K 2001 *Phys. Rev. E* **65**, 017106; Marsili M and Zhang Yi-Cheng 1998 *Phys. Rev. Lett.* **80**, 2741.
- [10] Axtell R L 2001 *Science* **293**, 1818.
- [11] Gabaix X 1999 *Quarterly Journal of Economics* **114**, 739.
- [12] Kechedzhi K E, Usatenko O V and Yampol'skii V A 2005 *Phys. Rev. E* **72**, 046138; Ree S 2006 *Phys. Rev. E* **73**, 026115.
- [13] Moreira A A, Paula D R, Costa Filho R N and Andrade J S 2006 *Phys. Rev. E* **73**, 065101(R).
- [14] Reed W J and Hughes B D 2002 *Phys. Rev. E* **66**, 067103.
- [15] Frieden B R, Plastino A, Plastino A R and Soffer B H 1999 *Phys. Rev. E* **60**, 48; 2002 *Phys. Rev. E* **66**, 046128.
- [16] Pennini F, Plastino A, Soffer B H and Vignat C 2009 *Phys. Lett. A* **373**, 817.
- [17] Frieden B R and Soffer B H 1995 *Phys. Rev. E* **52**, 2274; Frieden B R 1998 *Physics from Fisher Information*, 2nd Ed. (Cambridge Univ. Press, Cambridge); Frieden B R 2004 *Science from Fisher Information* (Cambridge Univ. Press, Cambridge).
- [18] Cohen-Tannoudji C, Diu B and Laloe F 2006 *Quantum Mechanics* (Wiley-Interscience, New York).
- [19] Pennini F and Plastino A 2006 *Phys. Lett. A* **349**, 15.
- [20] Frieden B R and Gatenby R A 2005 *Phys. Rev. E* **72**, 036101.
- [21] Zemansky M W and Dittmann R H 1981 *Heat and Thermodynamics* (McGraw-Hill, London).
- [22] Goldstein H, Poole C and Safko J 2002 *Classical Mechanics* 3rd Ed. (Addison Wesley, San Francisco).
- [23] National Statistics Institute, Spain, www.ine.es.
- [24] Ministry of the Interior, Spain, www.elecciones.mir.es
- [25] Electoral Commission, Government of the UK, www.electoralcommission.org.uk.
- [26] Hernando A, Villuendas D, Abad M and Vesperinas C (2009) arXiv:0905.3704v1 [cond-mat.stat-mech]

- [27] Census bureau website, Government of the USA, www.census.gov.
- [28] Journal Citation Reports (JCR) for 2007, Thomson Reuters
- [29] Limpert E, Stahel W and Abbt M 2001 *BioScience*, **51**, 341.
- [30] Fortunato S and Castellano C 2007 *Phys. Rev. Lett.*, **99**, 138701.
- [31] Gould H and Tobochnik J 1996 *An Introduction to Computer Simulation Methods: Applications to Physical Systems*, 2nd Ed. (Addison-Wesley).